

Assignment 4.

Vocabulary.

data , science , is , one , of , the
 most , important , courses , in ,
 computer , this , best , scientists
 perform analysis

Word	S1	S2	S3
data	1	1	2
science	2	1	0
is	1	1	0
one	1	1	0
of	1	1	0
the	1	1	1
most	1	0	0
important	1	0	0
courses	1	1	0
in	1	0	0
computer	1	0	0
this	0	1	0

best	0	1	0
best			
scientist	0	0	1
perform	0	0	1
analysis	0	0	1

V1 : { 1, 2, 1, 1, 1, 1, 1, 1 }

1, 1, 1, 0, 0, 0, 0, 0 }

V2 : { 1, 1, 1, 1, 1, 1, 1, 0 }

0, 1, 0, 0, 1, 1, 0, 0, 0 }

V3 : { 2, 0, 0, 0, 0, 0, 0, 0, 0 }

0, 0, 0, 0, 0, 0, 0 }

TF

word	s ₁	s ₂	s ₃
data	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{2}{6}$
science	$\frac{2}{12}$	$\frac{1}{9}$	0
is	$\frac{1}{12}$	$\frac{1}{9}$	0
one	$\frac{1}{12}$	$\frac{1}{9}$	0
of	$\frac{1}{12}$	$\frac{1}{9}$	0
the	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{1}{6}$
most	$\frac{1}{12}$	0	0
important	$\frac{1}{12}$	0	0
courses	$\frac{1}{12}$	$\frac{1}{9}$	0
in	$\frac{1}{12}$	0	0
computer	$\frac{1}{12}$	0	0
this	0	$\frac{1}{9}$	0
best	0	$\frac{1}{9}$	0
scientists	0	0	$\frac{1}{6}$
perform	0	0	$\frac{1}{6}$
analysis	0	0	$\frac{1}{6}$

IDF

$$\text{idf}(\text{data}) = \log(3/3) = 0$$

$$\text{idf}(\text{science}) = \log(3/2) = 0.18$$

$$\text{i,} = \log(3/2) = 0.18$$

$$\text{one} = \log(3/2) = 0.18$$

$$\text{of} = \log(3/2) = 0.18$$

$$\text{the} = \log(3/3) = 0.48$$

$$\text{most} = \log(3/1) = 0.48$$

$$\text{important} = \log(3/1) = 0.48$$

$$\text{courses} = \log(3/2) = 0.18$$

$$\text{in} = \log(3/1) = 0.48$$

$$\text{computer} = \log(3/1) = 0.48$$

$$\text{this} = \log(3/1) = 0.48$$

$$\text{best} = \log(3/1) = 0.48$$

$$\text{scientist} = \log(3/1) = 0.48$$

$$\text{perform} = \log(3/1) = 0.48$$

$$\text{analysis} = \log(3/2) = 0.18$$

$$IDF = \frac{1}{\log idf}$$

word	s1	s2	s3
data	0	0	0
science	0.03	0.02	0
is	0.015	0.02	0
one	0.015	0.02	0
of	0.015	0.02	0
the	0	0	0
most	0.04	0	0
important	0.04	0	0
courses	0.04	0.02	0
in	0.04	0	0
computes	0.04	0	0
this	0	0.53	0
best	0	0.053	0
scientist	0	0.053	0
perform	0	0	0
analysis	0	0	0